DOCUMENT RESUME

ED 344 906 TM 018 227

AUTHOR Moore, William P.
TITLE Testing Perceptions, Practices, and Malpractice: The Impact on Teachers of Court-Ordered Achievement Testing in a Desegregation Setting.
PUB DATE Apr 92
NOTE 58p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 1992).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Achievement Tests; Decision Making; Educational Practices; *Elementary School Teachers; Grade 3; Grade 4; Grade 5; Intermediate Grades; Mail Surveys; Primary Education; Questionnaires; *School Desegregation; School Districts; *Teacher Attitudes; Test Coaching; Testing Problems; Test Results; *Test Use
IDENTIFIERS *Iowa Tests of Basic Skills; *Mandated Tests; Teacher Assessment Practices Questionnaire

ABSTRACT

Teacher testing-related attitudes and practices were explored in the context of court-ordered achievement testing in a school district engaged in desegregation. Responses to a mail survey (the Teacher Assessment Practices Questionnaire) from 79 teachers (grades 3, 4, and 5) across 13 elementary schools in a midwestern urban district suggest that teachers engage in a large number of test preparation practices preceding mandated Iowa Tests of Basic Skills (ITBS) testing and use resulting test information for curricular and instructional decision making. Most teachers, however, report that the results of ITBS testing do not provide benefits that offset the time and costs associated with testing. Teachers also report having experienced pressure from others, inside and outside the school, to improve students' ITBS scores. Engagement in inappropriate, and at times unethical, testing practices was found to be substantially greater than had previously been reported. Unexpectedly, engagement in testing malpractice was found to be unrelated to the intensity of teachers' perceived pressure to improve test scores. Teachers who perceive greater value or benefit related to testing are less likely to engage in inappropriate testing practices. Thirteen tables present study findings, and there is a 24-item list of references. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED344906

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William P. Moore

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Testing Perceptions, Practices, and Malpractice: The Impact on Teachers of Court-Ordered Achievement Testing in a Desegregation Setting

William P. Moore

The University of Kansas

Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, April, 1992.

TM018227

BEST COPY AVAILABLE

ABSTRACT

The study had the goal of examining teacher testing-related attitudes and practices in a relatively unexplored educational and assessment setting: *court-ordered* achievement testing in a district engaged in desegregation. Responses to a mail survey instrument from 79 teachers (representing grades three, four, and five) across 13 elementary schools in a midwestern urban district suggested that teachers engaged in a large number of test preparation practices preceding mandated Iowa Tests of Basic Skills (ITBS) testing and used resulting test information for curricular and instructional decision-making. Alternately, most teachers reported that the results of ITBS testing did not provide benefits that offset the time and costs associated with testing. Furthermore, teachers reported having experienced pressure from others, inside and outside their school, to improve ITBS scores. Engagement in inappropriate, and often times unethical, testing practices was found to be substantially greater than had been previously reported in other studies of testing impact. Unexpectedly, engagement in testing malpractice was found to be unrelated to the intensity of perceived pressure to improve test scores teachers experienced. Additionally, as expected, teachers who perceived greater value/benefit of testing were less likely to engage in inappropriate testing practices.

Testing Perceptions, Practices, and Malpractice:
The Impact on Teachers of Court-Ordered Achievement Testing
in a Desegregation Setting

The implementation of a large-scale testing program, specifically set in place for district accountability or reform purposes, can bring with it unintended and abstruse consequences for those students, teachers, and administrators who must respond to the heightened interest in improved education by demonstrating increased test performance. The close association between testing and reform, particularly in the current restructuring reform effort, emphasizes an outcome-based orientation of educational effectiveness. While restructuring has focused on such important issues as enhancing teacher efficacy and empowering teachers in school and classroom-based decision-making, or creating innovative educational schools with the latest in instructional methods, technology and resources, the call for greater teacher and school accountability has been spirited. As expected, test score performance has attained a central role toward assuring that restructuring reform initiatives impact instruction and flow down to the student population. While mandated achievement examinations are not inherently iniquitous, there is great concern among psychometricians, measurement specialists, and testing directors that mandated achievement testing is not being locally accommodated nor implemented in a way that assures valid test score interpretation or generalization to a broader domain of cognitive skills (Bauer, 1990).

Current speculation of mandated testing impact paints a pressure-practice causal flow to explain recent revelations of teacher testing-related behavior. The pressure on teachers and administrators to improve test performance is theorized to

lead to unwanted instructional and curricular modification, teachers who focus upon the test content to the exclusion of other relevant instructional outcomes, and teacher malpractice and engagement in unethical test-related activities. As such, while pressure to improve performance is not necessarily a negative influence, it may well be the response to the pressure that can be counterproductive and potentially unethical.

Haladyna, Haas and Nolen (1990) have introduced the concept of test score pollution and indicate that "any factor which distorts a standardized test score interpretation can render the validity of such interpretation highly suspect" (p. 1). The authors suggested that divergent teacher practices (i.e., contrary to norming practices and test publisher guidelines) in preparation and/or administration of a standardized test is one form of test score pollution. According to other measurement specialists, psychometricians have reported a history of teacher testing practices characterized by 'misapplication and malpractice' (LeMahieu & Wallace, 1986). "Some estimates of the number of classrooms with exaggerated test scores directly influenced by teacher malpractice have been judged to be as high as 50%" (Perlman, 1985).

Haladyna, et al., (1991) clarified test score pollution by noting, " . . . pollution increases or decreases test performance without connection to the construct represented by the test, producing construct-irrelevant test score variance" (p. 4). Rightly so, the authors note the link to validity and the work of Messick (1984). Messick (1988), in what has become one of the most noted discussions of the dimensions of validity, explained " . . . what is to be validated is not the test or observation device as such but the inferences derived from test scores . . . inferences

about score meaning and interpretation and about the implications for action that the interpretation entails" (p. 13). As such, norm-referenced test score interpretations, such as those utilized in most school districts nationwide, can not be supported with an assumption of validity when divergent teacher practices in preparation and/or administration of the test are present. While this is not a new revelation, what is new are the increasing numbers of recent research efforts suggesting that our schools and teachers are not adhering to standard measurement practices.

Research has begun to examine the link between pressure to improve test scores and violations of the standardization assumption. Central to the recent exploration of testing impact has been a focus on test stakes (Corbett & Wilson, 1988). The research suggests that a test's stakes are related to the importance of decisions being made as a result of test performance and that decisions that affect future access can lead to pressure to perform well on the test. In such a high stakes test setting, teachers and administrators are each affected by student performance on the test if decisions about *their* effectiveness will be made on the basis of test scores. Low stakes testing programs are not immune to pressure nor to deviations from standardized testing practices if comparisons across classrooms, schools or programs are made or if media comparisons of school performance create the perceived need for improvements in achievement. Consequently, what seems unclear is whether pressure to improve scores alone contributes to differential teacher testing practices or whether other mechanism(s), perceptual or actual, operate that lead teachers to engage in practices which threaten the usefulness of the test score interpretation and cast doubt on the reliability of scores. It is parsimonious to conclude that those teachers who experience pressure to improve scores will engage

in practices that heighten the chances of higher scores. However, this has not been fully substantiated and was not found to be supported in the current study.

While the testing literature provides evidence of the effect of the imposition of state and locally mandated testing upon instructional and curricular decision making, the measurement and testing field is just beginning to examine the extent of the use of inappropriate assessment practices in the standardized testing environment. Additionally, lacking is an understanding of the specific types of preparation utilized by teachers and whether particular practices are preferred by teachers. Lastly, a greater understanding is needed of the impact of mandates and how, or even if, different testing mandate sources contribute to different testing perceptions and practices. This investigation of a mandated testing program in a large midwestern urban school district currently implementing a desegregation federal court order provides a unique opportunity to examine relationships among teacher testing-related perceptions and practices in an additional educational setting.

Background of Investigation

In the Kansas City, Missouri public school district a federal court order (Jenkins, et al. v. State of Missouri, et al., 1985) mandated the desegregation of the district. As part of the desegregation order, the court required the use of a standardized test (the district recommended the Iowa Tests of Basic Skills and the court approved its use) as a primary outcome measure of the effectiveness of the desegregation effort. ITBS testing is conducted annually at grades kindergarten through eight. TAP testing is also conducted at the high schools annually at each grade level. As such, each student is tested annually to provide information to the

court regarding desegregation effect on achievement. ITBS performance in the district has become a central feature of school effectiveness goals and planning efforts for district and school reform. There has been increasing interest (but not action as yet) by district leadership in using achievement scores to rank and evaluate schools, principals, and teachers as a measure of their educational effectiveness. The media regularly monitors achievement performance of the district and school-level comparisons are reported annually.

Assumptions of Investigation

While the current study explored testing impact in a district engaged in desegregation, the investigator makes no assumption that desegregation was the causative agent leading to the findings reported herein. Alternately, anecdotal reports to the investigator from teachers and principals within the district have suggested that pressure to improve test scores, as discussed later, stemmed from the court's and district administration's linkage of educational reform to performance on the mandated test. The study contributes to the growing body of evidence suggesting profound and consequential changes in instructional focus and delivery precisely because of the linkage between the imposition of a mandated testing program and the need to show substantial improvement in student test performance. The added dimension of court-ordered desegregation does not necessarily contribute to nor detract from the importance of this conclusion but may be a contributing factor through the heightened focus (by the court and by district leadership) on test score improvement in the district.

Procedures and Instrumentation

The Teacher Assessment Practices Questionnaire (TAPQ) used in the current study grew out of an effort at the Center for Educational Testing and Evaluation (CETE) at the University of Kansas to assess mandated testing impact. An instrument was constructed to gather data evaluating the impact of the Kansas Minimum Competency Testing (KMCT) program upon teacher attitudes and instructional practices (see Glasnapp, Poggio & Miller, 1991). The CETE effort stimulated the development of the TAPQ which is substantially different in scope and interest. The TAPQ was based on information obtained from a thorough review of the testing literature and discourse with testing professionals. The identification of potential test preparation practices was facilitated in part through the assistance of contacts with department of education personnel in 21 different states. As a result of these efforts 71 items were developed to measure teacher perceptions of the value and benefit of ITBS testing, and pressure to improve test scores (e.g., intensity and type). Testing-related actions of teachers were explored as well. The TAPQ queried participants about their use of resultant test information to make decisions about pupil performance and instructional and curricular decisions. Additionally, teachers were asked to indicate which, if any, of 36 testing-related practices spanning pre-assessment to post-assessment behavior were undertaken.

TAPQ Reliability

Because the data reported in this paper reflect just one aspect of a larger study exploring testing perceptions and preparation practices and achievement, in which the items were configured into a number of scales, individual item reliabilities are not reported (except where scales did not exist). However, scale reliabilities, of

which items are associated, are presented below. The discussion of results later will focus on individual item responses and relationships.

Reliability information based on coefficients of stability were gathered six weeks after return of the TAPQ (for a discussion of sample see Methods and Participants section). Approximately 20% of the respondents participated in the reliability effort (n=16). Stability indices (r_{xx}) are noted below in bold typeface. The following dimensions of a teacher's response to a mandated test were examined.

Pressure to improve test scores. This dimension expressed the extent to which teachers felt the existence of applied pressure. Pressure was defined as the recognition on the part of teachers that interested groups or individuals, whether internal or external to the district, were expecting improved student test performance. To the extent that pressure was felt by teachers, pressure was perceptually present. Five items measured a general perception of pressure (**r_{xx} = .79**):

(a) I feel there is pressure for this school to perform well on the ITBS.

(b) I feel pressure from others to raise math test scores on the ITBS,

(c) I feel pressure from others to raise reading test scores on the ITBS,

(d) I feel my principal has pressured me to spend increasing lengths of class time preparing my students for ITBS testing,

(e) I feel my school district is placing too much emphasis on ITBS performance as a measure of my instructional effectiveness.

Teachers responded to a four point Likert-type scale anchored at 1=strongly agree to 4=strongly disagree.

An additional two items measured the degree or intensity of pressure felt by teachers (Q64R-reading pressure and Q65M-math pressure). Teachers responded

to these items using a five point response scale anchored at the extreme options only. The scale ranged from "No Pressure"(coded 1) to "Extreme Pressure" (coded 5):

(Q64R) How much pressure was there on you as a teacher for maximizing or improving reading test scores on the last ITBS test administration? (r_{xx} = .92)

(Q65M) How much pressure was there on you as a teacher for maximizing or improving math test scores on the last ITBS test administration? (r_{xx} = .93)

Value of the testing program. The test information resulting from testing may be valuable to teachers for instructional decision-making, placement, and remediation. Value was defined as usefulness, contribution to a better understanding of student achievement and knowledge acquisition, and contribution to the teachers' efforts to create an effective instructional environment. As such, the value of testing was measured by five variables (r_{xx} = .70). Teachers responded using a four point Likert-type scale ranging from 1=strongly agree to 4=strongly disagree. The five items were:

(a) Overall, ITBS testing in the district has been beneficial for students.

(b) Overall, the ITBS tests provide me with information about my students that I generally cannot obtain from classroom assessments.

(c) Overall, the ITBS testing program has been valuable to me as a classroom teacher.

(d) Performance on the ITBS is an important indicator of the district's instructional effectiveness,

(e) In your opinion, the costs (time, money, loss of instructional time, etc.) of ITBS testing are justified given the information you get about your student's achievement and your instructional effectiveness.

Use of test information. Three dimensions of test information use were examined: (a) results used for pupil monitoring and feedback, (b) results used for instructional decision-making; and (c) results used for curricular decision-making. Teachers were asked to indicate whether a particular use for test information was undertaken during the current year using previous years' test information. Test information use for pupil monitoring and feedback was measured with 9 items. Test information use for curricular decision-making was measured with 4 items and instructional decision-making was measured with 5 items. Teachers responded either "Yes" (coded 1) or "No" (coded 0) to the following items:

The items used to assess uses for pupil monitoring and feedback ($r_{XX} = .87$) were:

(a) Informed your current students of their status on the ITBS based on last year's test results.

(b) Assessed whether pupils were meeting district performance expectations.

(c) Made reports to parents regarding their child's status on the ITBS.

(d) Diagnosed individual pupil's learning difficulties.

(e) Grouped pupils for instruction.

(f) Determined each pupil's instructional level.

(g) Assigned students for remedial instruction.

(h) Worked with counselors or school psychologists in decision-making about individual students based on ITBS scores.

(i) Identified students for enrichment-oriented instruction or activities.

For measuring instructional decision-making usage ($r_{XX} = .81$), the items were:

(a) Assessed your instructional effectiveness based upon last years test performance.

(b) Evaluated the adequacy of my instructional methods in basic skills areas.

(c) Matched instructional methods to classroom achievement levels.

(d) Studied the ITBS tests to assess their match to your instructional objectives.

Decisions regarding curricular modification were measured with the following items (r_{XX} = .84):

(a) Increased instructional emphasis in the existing curriculum on the material tested with the ITBS.

(b) Increased the amount of instructional time spent preparing students for the ITBS.

(c) Eliminated some curricular topics in order to spend more time upon ITBS-related topics.

(d) Revised the instructional scope and sequence in your class.

(e) Added lessons or units to the curriculum in your class.

As such, these dimensions of the use of test information attempted to assess the impact of the program through utilization of test results. It was assumed that teachers who had engaged in a number of these activities had, in fact, taken considerable efforts to use test information to assess instructional efforts and curricular alignment.

Testing practices and preparation. Among other practices, teachers may utilize a variety of test preparation strategies: test-taking skills training, testwiseness training, test familiarization and test practice activities, motivational activities, coaching strategies, or the utilization of unethical or inappropriate practices, that may improve test scores. Items asked teachers to indicate what practices they had utilized in their classroom during the test year with a view toward improved test performance. Teachers responded with a "Yes" or "No" indicating whether they

used the activity as a test preparation practice. Each preparation approach was defined through the identified practices associated with each strategy. The following types of test preparation approaches and practices were examined:

1. Test-taking Skills (r_{XX} = .68):

(a) Taught students how to follow test directions.

(b) Discussed with students the test-taking skills needed for the ITBS.

(c) Taught students time management skills for the ITBS.

(d) Discussed with students how to recheck their answers.

(e) Discussed with students how to mark the answer sheet correctly.

(f) Taught students how to answer multiple choice questions.

(g) Tried to reduce students anxiety about taking the test by giving them training in anxiety-reduction techniques.

2. Testwiseness Skills (r_{XX} = .82):

(a) Taught students some guessing strategies for multiple choice questions.

(b) Taught students some clues to help them find the correct answer.

(c) Gave students hints and strategies to help them do better on multiple choice items.

3. Test Familiarization Practices (r_{XX} = .88):

(a) Provided class with practice questions like those found on the test.

(b) Showed past versions of the ITBS so they would know what to expect.

(c) Used other tests of similar format to practice the procedures of the test.

4. Motivational Practices (r_{XX} = .70):

(a) Had contests before the test to motivate pupils to do well on the ITBS.

(b) Provided prizes and incentives to pupils who worked hard preparing for the ITBS.

(c) Tried to create an exciting classroom atmosphere around the test period by using signs, posters, and other spirit-related activities.

(d) Developed a positive attendance week around the testing week and provided rewards for high attendance.

(e) Scheduled a fun activity at the conclusion of the testing program.

5. Inappropriate Testing Practices (r_{XX} = .43):

(a) Praised those students who answered an item correct during testing.

(b) Recoded a student's answer on the answer sheet because I knew they just miscoded their response.

(c) Called my best students aside and encouraged them to perform well.

(d) Gave students practice exercises using questions off the current test.

(e) Used prior years test items as practice items for this years test.

(f) Taught students how to approach a question I thought would be on the test.

(g) During the test administration I tried to help a student better understand how to answer a question by giving hints and clues to the correct answer.

(h) Assigned test-related homework for students to complete during vacations or on weekends.

(i) Modified the time schedule during testing to accommodate our class schedule.

(j) Gave students practice on items which were directly off the test, with a change in the stem or distractors.

(k) Gave students as much time as they needed to make sure every student completed the test.

6. Intensive Coaching or Practice (r_{xx} = .61):

(a) Conducted special reviews or content drills in preparation for testing.

(b) Focused upon extensive drill and practice on skills or items similar to those tested.

(c) Taught students how to use deductive reasoning skills to identify the correct answer.

(d) Used a commercially developed test preparation package such as *Scoring High on the ITBS*.

A final series of items were used to query teachers about their general approach to test preparation in their classroom during the test year. Using three general types of test preparation as suggested by Anastasi (1981) teachers were asked to choose the one **general** approach they relied upon most often while preparing their students for the ITBS reading and math tests. The strategies were:

(a) A short test-taking orientation and a brief practice session(s).

(b) A long, intensive coaching program that includes concentrated drill and practice on a sample of items or skills similar to those tested on the ITBS.

(c) A broad-based approach with instruction to improve many skills but not necessarily any one skill tested on the ITBS.

(d) I did not provide any test preparation for my students on the ITBS, other than day-to-day instruction for the regular curriculum.

The reliability coefficient for general preparation approach for reading was r_{xx} = .94 and for mathematics preparation, r_{xx} = .96.

Method and Participants

The questionnaire instrument was mailed to 166 (third through fifth grade) teachers in a random sample (n=29) of 51 elementary schools servicing third through fifth grade students in the Kansas City, Missouri School District. This

represented all third, fourth, and fifth grade classroom teachers in the sampled schools. The population of district-wide third through fifth grade teachers was N=398. As such, the proposed sample represented 42% of the teaching population. The proposed sample of schools represented 57% of all elementary schools with grades three through five.

All elementary schools in the district were expected to demonstrate improved achievement from prior years and each school received equivalent funding to facilitate achievement growth in basic skills through an effective schools grant to improve reading and mathematics achievement.

The instrument packet was timed to arrive at the teacher's school prior to summer recess and about 4 weeks after testing was completed for the 1989-90 academic year. The instrument packet included an incentive offer response card and pre-addressed and stamped envelope to return the completed questionnaire and incentive response card which allowed respondents to receive $2.00 for their participation. Teachers were asked to reflect upon their perceptions and practices related to ITBS testing for the just completed test year (1989-1990).

Results

TAPQ Response Rate

With an initial mailing and two follow-up mailings, 78 teachers responded. In an alternate effort to elicit participation, a random sample (17 teachers; 20%) of the remaining 88 non-respondents were contacted by telephone for participation. This effort produced one respondent willing to participate. The final sample

consisted of 79 teachers for a response rate of 48%. Grade three teachers were less willing than fifth grade teachers to participate as their response rate was 38%. Forty-eight percent of the fifth grade teachers sampled responded to the instrument. Grade four teachers, unexpectedly, were not well represented in the study (10 participants). It is unclear why this rate of participation deviated from that of the other grade levels sampled. Of the 29 schools sampled, 13 were represented in the final sample for an inclusion rate of 45%. Of the population of schools, 25% (i.e., 13 of 51) of the schools were represented. Of the population of teachers, 20% (i.e., 79 of 398) were represented in the participant sample.

Testing Perceptions and Practices

The 71 item instrument was examined, both aggregated and disaggregated by teaching assignment, for the rate of affirmative responding by item (see Tables 1 to 10). A summary of comparable research and the results of the current study follows.

Use of test results for pupil monitoring. The use of test performance information for pupil monitoring and feedback was examined with 9 items (see Table 1). Teachers were asked if they had engaged in specified practices during the just completed academic year using *prior year* (1988-89) test result information.

The most common use reported by teachers was that associated with determining individual and group standing relative to district expectations. Ninety-three percent of the teachers reported using test performance information to assess whether the pupil met district performance expectations. Eighty-two percent of the teachers made reports to parents regarding their child's status on the ITBS. Alternately, only 62% informed their students of their status on the ITBS based on

prior year's results. Similarly, 63% used test results to determine each pupil's instructional level and 61% used test results to group pupils for instruction. Slightly more than 70% used the ITBS results to diagnose individual pupil learning difficulties and 68% used the test results to identify remedial instruction efforts.

Insert Table 1 about here

Use of test results for instructional and curricular decision-making. The ITBS appeared to have a profound effect upon instructional and curricular decision-making. For the 4 years preceding testing as well as the current test year, teachers were asked to indicate whether they had used test performance information during that time to make decisions regarding their instructional effectiveness and for curricular revisions. Greater than 90% of the teachers had increased the instructional emphasis of the material tested on the ITBS in the existing curriculum (97%) and 89% had added lessons or units to the curriculum they taught (see Table 2). Furthermore, 74% had eliminated some curricular topics in the existing curriculum in order to spend more time on ITBS-related topics or skills.

The curricular and developmental sequencing of the topics tested appeared to have influenced teachers as well. Eighty-eight percent had revised the instructional scope or sequence in their classes and 88% had studied the ITBS tests to assess their match to their instructional objectives. Test performance information was used to a great extent to assess teaching effectiveness (91%) and to evaluate the adequacy of instructional methods in the basic skills area (94%).

Insert Table 2 about here

The modification of instruction and curriculum to more closely reflect the emphasis of test content has been noted in other studies (Wood, 1988; Rottenberg and Smith, 1990; Nolen, Haladyna, & Haas, 1990; Popham, Cruse, Rankin, Sandifer, & Williams, 1985; Bateson, 1990). Specifically, as Madaus (1987) has suggested, a two-tiered curriculum evolves with the implementation of a mandated testing program.

Fish and Allard (1990), in an ethnographic study of 36 elementary teachers in two California school districts with mandated standardized testing serving an accountability function (assessing educational quality), found teachers reporting the perception that curricular changes made in response to the test had reduced the quality of their instruction. According to Fish and Allard (1990), 50% of the teachers perceived the existence of two different curricula: test preparation and content, and the traditional regular curriculum.

Engagement in testing practices and preparation. Ligon and Jones (1982) defined test preparation for standardized testing as, "One which contributes to students performing on the test near their true achievement levels, and one which contributes more to their scores than would an equal amount of regular classroom instruction" (p.1). Matter (1986) suggested that test preparation activities should not be additional activities undertaken by the teacher but should be incorporated into the regular ongoing instruction as often as possible.

In the current study it was found that teachers reported high rates of utilization for most of the preparation activities examined (see Tables 3 through 8). In particular, all teachers reported teaching students to follow test directions, and encouraging students to eat a good breakfast, get enough sleep, and be rested for the days of testing (see Table 3). Greater than 90% of the teachers said they taught students how to use reasoning skills to identify the correct answer (96%); discussed with students how to mark the answer sheet correctly (94%); and discussed with students the test-taking skills needed for the ITBS (96%). Unexpectedly, 93% of the teachers reported utilizing some form of test preparation package. Anecdotal evidence obtained from teachers during and after the test year indicated that many teachers used *Scoring High on the ITBS.*

Similarly, Fish and Allard (1990) found teachers had increased efforts to teach to the test and use commercial test preparation packages. Furthermore, the test influenced teachers to modify the type of assignments given to students. They reported increased use of multiple-choice types of activities and assessment formats.

Insert Table 3 about here

Teachers in the current study, relative to other preparation practices, reported less utilization of motivational strategies to prepare students for ITBS testing. Only one activity was found to be a much utilized practice. Eighty-two percent indicated providing students with a "fun activity" at the conclusion of testing. Alternately, approximately fifty percent of the teachers used prizes or other incentives for

preparing hard for the ITBS (Q21). Forty-two percent had class contests to motivate students before testing (Q43). Other practices are reported in Table 4.

Insert Table 4 about here

Teachers in the current study were prolific providers of testing practice and guessing strategies. Ninety percent had provided practice on tests of a similar format to practice the format and procedures of the ITBS tests. Furthermore, 83% taught students how to answer multiple-choice items and discussed with students how to recheck their answers. Approximately 80% taught students guessing strategies and taught students time management skills for the ITBS (84%).

Insert Table 5 about here

Two items asked teachers to indicate their general or global approach to test preparation in reading and mathematics content. Based on Anastasi's descriptions of test preparation, teachers were given four options including, "Did not provide any preparation other than day-to-day instruction for the regular curriculum"(see Tables 6 and 7).

Insert Tables 6 & 7 about here

Obtained reliability coefficients and the relative similarity of response patterns across the two content areas suggested that a teacher's general approach to preparation was a relatively static behavior and did not vary substantially across content or time. Teacher responses indicated that only 3% of the teachers did not use some form of systematic preparation for mathematics content. Similarly, only 6% did not engage in systematic preparation for reading content. The majority of teachers preferred a broad-based approach with instruction to improve many skills but not necessarily any one skill tested on the ITBS. Approximately, one-quarter used a long, intensive coaching program with concentrated drill and practice on skills similar to those tested. Slightly less than 20% used a short test-taking orientation and brief practice sessions.

Inappropriate testing practices. Of particular concern to measurement professionals is whether teacher and district testing practices and preparation violate test standardization principles. A valuable product of the current study was the self-reported rate of teacher engagement in practices generally considered inappropriate and potentially unethical. While other studies have tried to observe these behaviors in classroom settings, and others have asked teachers whether they would consider a behavior inappropriate, the current study was decidedly more direct. Teachers were simply asked if they had engaged in a particular behavior in their efforts to prepare their students for ITBS testing. The instrument did not identify inappropriate practices, nor were teachers asked whether they believed the practices were inappropriate. The anonymity of the approach (e.g., mail survey without demographics) may have elicited the surprising incidence rates, whereas direct

observation of a teacher may have a suppression effect upon the rate of observed behaviors, especially potentially unethical practices.

Haladyna, Haas, and Nolen (1990) have proposed a continuum of test preparation activities anchored with ethical to highly unethical behaviors. At the ethical end of the continuum were placed such activities as: (a) training in testwiseness skills, (b) checking answer sheets, and (c) increasing student motivation through appeals to parents, students, and teachers. Unethical activities mentioned were: (a) developing a curriculum based on test content, (b) preparing objectives based on items on the test and teaching accordingly, (c) presenting items similar to those on the test, and (d) using score-boosting commercial test preparation packages. Highly unethical activities noted were: (a) dismissing low-achieving students from testing, and (b) pretesting items verbatim from the test to be given (p. 33).

Mehrens and Kaminski (1989) have suggested an alternative continuum of test preparation practices:

1. General instruction on objectives not determined by looking at the objectives measured on standardized tests (ethical);
2. teaching test taking skills (typically considered ethical);
3. instruction on objectives generated by a commercial organization where the objectives may have been determined by looking at objectives measured by a variety of standardized tests;
4. instruction based on objectives (skills, subskills) that specifically match those on the standardized test to be administered;

5. instruction on specifically matched objectives (skills, subskills) where the practice (instruction) follows the same format as the test questions;

6. practice (instruction) on a published parallel form of the same test (unethical);

7. practice (instruction) on the same test as that administered (unethical).

Shepard and Kreitzer (1987) suggested that a grey area exists in test-related instruction in which appropriate forms of test preparation can potentially become unethical. Mehrens and Kaminski (1989) noted that this point exists somewhere between items 3 and 5 above and the point is determined by the scope of inference the teacher wishes to make from the test scores (p. 16). Given the norm-referenced interpretation of the ITBS, and the test development process which assumes that items are representative of a broader domain of skills, the generalizability and potential for inference of virtually any scope would be hindered if inappropriate practices, such as those found in the study, were utilized.

Alternately, there is considerable debate about what constitutes unethical and inappropriate practices. For the purpose of this study, inappropriate practices were identified as such if utilization would violate standardization assumptions or would artificially elevate student scores through some teacher intervention after testing had occurred. In some instances, practices were consistently identified as unethical in the testing literature (e.g., teaching current year test questions as practice exercises) and were included.

At the same time teachers are reporting extensive use of a variety of acceptable preparation practices, they are also indicating engagement in unsound testing practices (see Table 8). Given the incidence rates found for items measuring

potentially unethical and/or inappropriate testing-related practices, and the level taught (elementary school) the findings discussed next may need a caveat: elementary school students are generally more dependent upon their teachers during testing given their relative inexperience with standardized tests and optical-scan answer sheets. As such, teachers may be more active during testing to quell fears, address test anxiety, and provide support. This behavior may well stimulate teachers to behave in ways that in retrospect would be considered inappropriate but at the time were necessary for student emotional support. Furthermore, the rate of responding and the apparent willingness to provide information regarding potentially unethical activities may suggest that these teachers were not informed about what may or may not be appropriate assessment-related activities.

This important dimension, teacher intent and understanding of appropriate practices, was not examined in the current study . The question of whether teachers engage in specific practices intending to improve test scores with the knowledge that the practices were indeed inappropriate is unclear. Research currently being conducted in the same district *tentatively* suggests that for many of the practices examined in this study, teachers are not able to differentiate between appropriate and inappropriate testing practice (Moore, 1992).

Based on the findings of Gonzalez (1985), in a study of teacher perceptions of test-related cheating, it may well be accurate to assume that the teachers examined in this study were largely uninformed regarding appropriate and inappropriate forms of testing practice. Gonzalez's results suggested that considerable disagreement existed among educators regarding practices considered inappropriate. Regardless, the rates of utilization found were unexpected and compared with other studies of

testing impact suggest a much greater problem in the elementary schools in the district examined than has been found in other studies of testing impact in elementary schools.

Somewhat less than half of the sampled teachers reported using prior year's test questions as practice for the current year test (44%). Further, 17% of the teachers said they gave students practice exercises using questions directly off the current year's test. Similarly, 24% gave students practice on questions which were directly off the current test with a change in the stem or distractors. Sixty percent of the sample indicated they had taught their students how to approach an item they thought would be on the test. Lastly, 26% of the teachers indicated they had recoded a student's answer sheet because they knew that the student(s) had just miscoded the answer.

Nolen et al's. (1990) findings of teacher initiated test preparation practices appear to support these findings but in may cases the rates of engagement differ widely, with the current study teachers exhibiting substantially more usage of inappropriate practices. Nolen, et al., found that 25% of their sample taught vocabulary items covered on the test; 11% provided practice on items from the prior year's test and 10% reported teaching students the actual test item off the current year's test (p.10). Further, 20% provided students with additional examples during the test administration and 10% gave rewards for completing the test.

Insert Table 8 about here

Other items measuring inappropriate testing-related activities also were examined in the current study (see Table 8). Twenty-four percent of the teachers indicated that they extended the testing time limits beyond those suggested by the test publisher to give students as much time as they needed to make sure every student got to complete the test. In addition, 41% of the teachers modified the testing schedule (shortened or lengthened) to better accommodate their instructional schedule. Twenty-three percent provided hints and clues during the test to help a student or students better understand how to answer a question. Furthermore, 23% praised a student or students, during the test administration, who answered a question correctly. Lastly, 18% said they had called their best students aside, prior to testing, and encouraged them to do their best on the test.

Perceived pressure to improve test scores. Teachers were, to a large extent, reporting moderate to extreme perceived pressure from superiors and others within the district to improve or maximize student test scores (see Table 9). Almost half of the teachers had been encouraged by their superiors to improve test scores using "any method available." Similarly, 58% of the respondents felt that their principal had pressured them to spend increasing lengths of class time preparing their students for ITBS testing. Eighty-five percent of the teachers believed there was too much emphasis upon test scores in their school and greater than sixty percent had experienced pressure from others to raise math and reading test scores.

Insert Table 9 about here

Teachers were asked to indicate **how much** pressure they experienced to raise student test scores on the ITBS (Q64R and Q65M). The mean response on a scale of 1=No pressure to 5=Extreme Pressure, was 2.6 for both reading and math tests indicating a moderate degree of pressure was experienced (see Table 8). Eight percent of the sample reported extreme pressure to improve math and reading test scores. Fifty-six percent reported levels of pressure ranging from moderate to extreme for math improvement and 53% reported the same for reading.

Perceived value of the testing program. Teachers were asked to rate their level of agreement (1= strongly agree to 4=strongly disagree) with statements regarding the benefits of ITBS testing in their schools. For ease of presentation and interpretation items have been rescaled to 1=strongly disagree to 4=strongly agree (see Table 10). However, in analyses, items retained their original scaling noted above. Typically, the mean response favored the disagree position. Teachers did not perceive the testing program to be valuable to them as classroom teachers (mean= 2.41; 65% did not find the program valuable); or to be beneficial for students (mean= 1.97; 75% did not believe it to benefit student learning); or to provide them with information about their students not generally obtained from classroom assessments (mean= 1.94; 75% did not get unique information from the tests).

Insert Table 10 about here

Teachers in the current study generally did not believe that classroom test performance on the ITBS was an important indicator of their instructional

effectiveness (mean= 1.89; 76% did not believe the test results to be an important indicator). Lastly, greater than 80% did not believe that the testing program was worth the time or costs sacrificed given the results gained regarding learning and teaching (mean=1.87).

Other studies of testing impact support the findings reported above. Haas, Haladyna, and Nolen (1989) reported that greater than 80% of the teachers and administrators thought that the state-mandated test was not worth the time and money spent.

Rottenberg and Smith (1990), in an examination of the impact of external testing upon teacher attitudes and behaviors, noted that external testing reduced the teachers' sense of efficacy and competence.

Relationships Among Testing Perceptions and Practices

In order to explore the relationships among testing practices such as engagement in appropriate and inappropriate test preparation and perceptions of pressure and value zero-order correlations were examined (see Tables 11 through 13).

Pupil monitoring and feedback. The results suggested that teachers who perceived greater pressure to improve test scores tended to be less likely to use test performance information to identify enrichment opportunities (Q9) for students ($r_{q9\ q64} = -.32$, $p \leq .01$; $r_{q9\ q65} = -.33$, $p \leq .01$). As such, enrichment opportunities for students appear to suffer when the teacher perceives greater pressure to improve test scores. Otherwise, there were no significant relationships between the degree of perceived pressure and use of test information for pupil monitoring and feedback (see Table 11).

Similarly, there were few significant relationships found between perceived value of the test program and use of test information for pupil monitoring and feedback (see Table 11). An exception was found for the relationship between use of test information for remedial instruction (Q8) and perceived value of the program for providing information not generally obtained from classroom assessments (Q68). Teachers tended to find greater value when testing helped identify remedial opportunities ($r_{q8\ q68}$ = -.24, p≤.05; Q68 scaled 1=strongly agree to 4=strongly disagree). Additionally, teachers who had used test information while working with counselors and school psychologists about a child (Q10) tended to be less sure the testing program benefited students, or that the test was an important indicator of their instructional effectiveness, or that the costs associated with testing were justified ($r_{q10\ q67}$ = .30, p≤.05; $r_{q10\ q69}$ = .27, p≤.05; $r_{q10\ q70}$ = .29, p≤.05).

Insert Table 11 about here

Instructional decision-making. Teachers who reported greater intensity of pressure to improve test scores tended to use test information to assess their teaching effectiveness (Q12) ($r_{q12\ q64}$ =.27, p≤.05 ; $r_{q12\ q65}$ = .24, p≤.05) and to match instructional methods with student ability levels (Q13) ($r_{q13\ q64}$ =.27, p≤.05; $r_{q13\ q65}$ = .25, p≤.05) (see Table 12). Generally, however, use of test results for instructional decision-making was found to be unrelated to the perceived value and benefits of the testing program.

Insert Table 12 about here

Curricular decision-making. The tendency to make curricular modifications such as revising the scope and sequence of curriculum (Q14) was found to be significantly related to the intensity of perceived pressure to improve test scores in reading and math ($r_{q14\ q64}$ = .27, p≤.05: $r_{q14\ q65}$ =.29, p≤.05). As such, teachers who perceived more intense pressure to improve scores also tended to revise the scope and sequence of their curriculum. Furthermore, teachers who tended to make curricular revisions (Q14) also tended to perceive less benefit for students from the testing program ($r_{q14\ q67}$ =.26, p≤.05). However, in all other aspects examined, perceived value was unrelated to curricular decision-making.

Inappropriate testing practices. The results suggest no significant relationship between the perceived intensity of pressure to improve test scores and engagement in inappropriate testing practices (see Table 13). Alternately, significant relationships were found between the degree of perceived value of the testing program and engagement in inappropriate testing practices. Interestingly, the only significant relationships found between the perceived value of the test and appropriate and inappropriate testing practices were those in which greater utilization of inappropriate practices were related to lower perceived value. As such, teachers who engaged in inappropriate testing practices tended to have poorer perceptions of the value of the testing program in terms of its benefit to themselves and to their students. For example, teachers who indicated recoding student(s) answer sheets after testing (Q27) tended to perceive the program as: (a) not valuable to classroom

teachers, (b) not providing unique/valuable information about students, and (c) not providing a good indication of their instructional effectiveness.

Insert Table 13 about here

Summary

Teachers in the current study were strongly influenced by the test to modify their instructional efforts and curriculum. Greater than 90% added test-related lessons to the curriculum and more than 70% eliminated topics in order to spend more time on test-related skills. Furthermore, almost all teachers had increased the emphasis of test-related content/skills in their instructional efforts and increased the amount of instructional time spent preparing for the tests.

The current study found that greater than 90% of the sample used some form of systematic test preparation strategy. Surprisingly, the chosen strategy did not differ substantially across reading and mathematics curriculum. This would suggest that the sample teachers did not generally discriminate between curricular areas when preparing students for testing.

Teachers did not typically consider the testing program to be of benefit or have value in their instructional efforts, nor in terms of student learning. Furthermore, teachers felt they could get the same or better achievement information from their own classroom assessment results. Additionally, teachers generally felt that the test was <u>not</u> a good indicator of their instructional effectiveness but did utilize test result information to make instructional and curricular decisions.

The study has confirmed the existence of pressure to improve test scores in an additional setting. Responses indicated that teachers had typically been encouraged by supervisors to improve test scores "using any method available." Teachers felt pressure from their principal to spend greater time preparing students for testing. Furthermore, teachers perceived greater pressure to improve math test scores than they did reading test scores. Lastly, almost half of the participants were aware of others in their school who had engaged in "cheating" to improve student test scores.

The current study found substantially greater engagement rates in inappropriate and potentially unethical forms of testing behavior than had been reported in studies elsewhere. However, the utilization of appropriate and inappropriate testing practices was found to be unrelated to the perception of pressure or degree of pressure teachers reported.

Alternately, teacher perceptions of value and benefit were found to be significantly related to the utilization of inappropriate testing practices. Overall, these results do not support the common belief that teachers under pressure to improve test scores will resort to inappropriate (or for that matter appropriate) testing practices to demonstrate improved student performance. While teachers in this sample did report moderate to extreme pressure to improve scores and did engage in inappropriate practices, the presumed relationship was not supported. Consequently, the question of what led teachers to use these practices can be in part explained by the relationships found between value variables and utilization of inappropriate practices. Lower perceived value was related to a greater utilization of these practices. Of course causation cannot be implied and begs the question of whether

utilization of inappropriate practices leads teachers to feel less value for the program or whether the perception of less value leads teachers to use these practices.

Implications

Overall, the current study has identified the existence of pressure to improve test scores in an additional setting, typically considered by most conventions to be a low-stakes setting. No graduation requirements are attached to the test, teachers are not formally evaluated based on student achievement, and no formal consequences are incurred by the teachers or students when achievement is substandard. Alternately, teachers experienced pressure and have reported that pressure is being directed at them to improve scores.

The evidence that teachers are dissatisfied with this climate of pressure and accountability appears to be supported through responses to pressure, value and benefit questionnaire items. Further, it appears that teachers, by indicating that the test does not accurately reflect their instructional effectiveness, are suggesting that the utilization of a test constructed for nationwide use will not effectively address local educational needs and outcomes.

The results of this study appear to speak to the changing role of teachers and testing in the current educational climate. The progressive change in the nature of standardized testing from a diagnostic/instructional tool to one centered upon evaluation and accountability has teachers scrambling to address a growing desire of varied stakeholders, policy-makers, and consumers of education for demonstrated improvement in the quality of education. The substantial reported use of inappropriate testing practices reported in this study supports the growing body of

evidence suggesting that teachers are discarding their traditional role of classroom decision-maker and instructional leader to one of a passive recipient of directives and mandates to "correct the situation." The need to show demonstrated achievement growth has created a climate oriented toward efficiency. Teachers respond with the 'quick fix' (e.g., aligned curriculum, test preparation) which in some cases appears to have produced the desired outcome: higher test scores. What appears to be lacking is an educational climate focused toward improved learning instead of improved test performance. Certainly, the former promulgates the latter, but the immediate pressure for test score improvement encourages a reconstruction of the curriculum to meet the content of the basic skills test and not the content perceived by the former instructional leaders to be most important for their students. For some teachers, the data suggest, this may mean using practices of an unethical and/or inappropriate nature.

Standardized testing has held a central role in education for many years. What studies of testing program impact have most recently demonstrated is the growing reliance on test scores for decision-making and the increasing potential for misuse of test scores. Educational and political policy-makers need to address the important link between instruction and testing and ensure that teachers are integrated into, not isolated from, the intent of testing. These data suggests a need exists to better prepare teachers to integrate standardized testing in the classroom; so to better serve instructional decisions and to minimize the disruption of the leadership roles teachers have historically held in their own classrooms. While testing continues to be a central feature of educational reform efforts, the need for systematic, well conceptualized integration appears lacking. If standardized testing is ever going to

be positively and beneficially integrated into the instructional environment, the key to that effort will be the inclusion of teachers into the decision-making process regarding assessment purposes and priorities, as well as a better understanding on the part of policy-makers of the appropriate role of achievement testing in our nation's schools. As Fish and Allard (1990) note, ". . . procedures for evaluation will fail if they are used to measure outcomes without attention to the change process and the meaningful connection of curriculum and instruction to assessment" (p. 19). Otherwise, if test results continue to be the central measure of district, school, or teacher effectiveness, not only will reformers be ignoring other relevant outcomes, but the critical ingredient in the process of reform will be isolated and forced to assume a passive recipient, instead of active participant, role in the effort to bring about meaningful and long-term changes in education.

References

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. American Psychologist, 36, 1086-1093.

Bateson, D. J. (1990). What are some of the effects of province-wide examinations: The findings. Paper presented at the annual joint meeting of the National Council on Measurement in Education and the American Educational Research Association, Boston.

Bauer, E. (1990). Testing practices and issues: NATD survey results. Oakland, CA: Oakland Public Schools.

Corbett, H. D., & Wilson, B. (1988). Raising the stakes in state-wide mandatory minimum competency testing. Politics of Education Association Yearbook 1988, 27-39.

Fish, J., & Allard, J. (1990). The impact of mandated standardized testing. Paper presented at the annual meeting of the American Educational Research Association, Division B, Boston: MA.

Glasnapp, D. R., Poggio, J. P., & Miller, M. D. (1991). Impact of a "low stakes" state minimum competency testing program on policy, attitudes, and achievement. Advances in Program Evaluation, 1B, 101-140.

Gonzalez, M. (1985). Cheating on standardized tests: What is it? In P. Wolmut, & G. Iverson (Ed.), National association of test directors 1985 symposia (pp. 4-16). Portland, OR: Multnomah ESD.

Haas, N., Haladyna, T. M., & Nolen, S. B. (1989). Technical Report 89-3: Standardized testing in Arizona: Interviews and written comments from teachers and administrators. Phoenix, AZ: Arizona State University.

Haladyna, T., Haas, N., & Nolen, S. (1990) Test score pollution. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Jenkins, et al, v. State of Missouri, et al. 639 F. Supp. 19 (W. D. Mo. 1985).

LeMahieu, P., & Wallace, R. (1986). Up against the wall: Psychometrics meets praxis. Educational Measurement: Issues and Practice. 5(1), 12-16.

Ligon, G. D., & Jones, P. (1982). Preparing students for standardized testing: One district's perspective. Paper presented at the annual meeting of the American Educational Research Association, New York.

Madaus, G. F. (1987). Testing and the curriculum. Chestnut Hill, MA: Boston College.

Matter, M. K. (1986). Legitimate ways to prepare students for testing: Being up front to protect your behind. In J. Hall, & P. Wolmut (Ed.), National Association of Test Directors 1986 Symposia (pp. 10-11). Oklahoma City, OK: Oklahoma City Public Schools.

Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? Educational Measurement: Issues and Practice, 8(1), 14-22.

Messick, S. (1984). The psychology of educational measurement. Journal of Educational Measurement, 21, 215-237.

Messick, S. (1988). Validity. In R. L. Linn (Ed.) Educational Measurement. Washington, DC: American Council on Education.

Moore, W. P. (1992). Testing Practices and Preparation: Teacher differentiation between appropriate and inappropriate practices. (Work in progress).

Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1990). A survey of actual and perceived uses, test preparation activities, and effects of standardized achievement tests. Paper presented at the joint annual meetings of the American Educational Research Association and the National Council for Measurement in Education, Boston.

Popham, W. J., Cruse, K. L., Rankin, S. C., Sandifer, P. D., & Williams, P. L. (1985). Measurement-driven instruction: It's on the road. Phi Delta Kappan, 66, 628-634.

Perlman, C. (1985). Results of a city-wide testing program audit. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Rottenberg, C., & Smith, M. L. (1990). Unintended effects of external testing in elementary schools. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Shepard, L. A., & Kreitzer, A. E. (1987). The Texas Teacher Test. Educational Researcher, 16(6), 22-31.

Wood, T. (1988). State-mandated accountability as a constraint on teaching and learning science. Journal of Research in Science Teaching, 25(8), 631-647.

Table 1

Proportion (sd) of Teachers Engaging in Pupil Monitoring and Feedback as a Use for Test Information

	Teaching Assignment-Grade			
Item number and question	3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
2. Informed students of status	.53 (.51)	.60 (.52)	.77 (.43)	.62 (.49)
3. Reports to parents of status	.84 (.37)	.70 (.48)	.82 (.39)	.82 (.39)
4. Grouped pupils for instruction	.63 (.49)	.40 (.52)	.68 (.48)	.61 (.49)
5. Diagnosed pupil learning	.74 (.45)	.60 (.52)	.73 (.46)	.72 (.45)
6. Determined pupil instructional level	.71 (.46)	.40 (.52)	.61 (.50)	.62 (.49)
7. Assessed pupil standing relative to district expectations.	.89 (.31)	1.00 (.00)	.95 (.21)	.93 (.26)
8. Assigned students for remedial instruction	.74 (.45)	.40 (.52)	.68 (.48)	.68 (.47)
9. Identified students for enrichment	.84 (.37)	.50 (.53)	.77 (.43)	.77 (.42)
10. Worked with counselors regarding pupil	.61 (.50)	.80 (.42)	.62 (.50)	.64 (.48)

Note: Figures reported have been rounded to two decimal places.

Table 2

Proportion (sd) of Teachers Engaging in Instructional and Curricular Decision-Making using Test Information

	Teaching Assignment-Grade			
Decision Area Item number and question	3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
Instructional Decision-Making				
11. Evaluated Instructional methods	.92 (.27)	.90 (.32)	1.00 (.43)	.94 (.49)
12. Assessed teaching effectiveness	.87 (.34)	1.00 (.00)	.95 (.22)	.91 (.28)
13. Matched instructional methods to ability levels	.87 (.34)	.90 (.32)	.95 (.21)	.90 (.30)
17. Studied tests to assess match to instructional objectives	.86 (.35)	1.00 (.00)	.90 (.30)	.88 (.32)
Curricular Decision-Making				
14. Revised curricular scope & sequence	.86 (.35)	.70 (.48)	1.00 (.00)	.88 (.32)
15. Added lessons or units	.87 (.34)	.90 (.32)	.96 (.21)	.89 (.32)
16. Increased emphasis of ITBS material	.95 (.23)	1.00 (.00)	1.00 (.00)	.97 (.17)
18. Increased coverage time of ITBS test material	.89 (.31)	.90 (.32)	.91 (.29)	.90 (.30)
19. Eliminated some topics to spend more time on ITBS content	.68 (.47)	.80 (.42)	.82 (.39)	.75 (.44)

Note: Figures are rounded to two decimal places.

Table 3

Proportion (sd) of Teachers Engaging in Test-taking and Testwiseness Training in Preparation for ITBS Testing

	Teaching Assignment-Grade			
Practices Item number and question	3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
Test-taking Skills				
20. How to follow test directions	1.00 (.00)	1.00 (.00)	1.00 (.00)	1.00 (.00)
23. How to mark answer sheet	.97 (.16)	1.00 (.00)	.86 (.35)	.94 (.23)
32. Test-taking skills needed	.95 (.23)	1.00 (.00)	.95 (.21)	.96 (.20)
37. Anxiety-reduction techniques	.47 (.51)	.50 (.53)	.64 (.49)	.52 (.50)
39. Time management skills	.82 (.39)	.80 (.42)	.91 (.29)	.85 (.36)
41. How to answer multiple-choice items	.79 (.41)	.90 (.32)	.91 (.29)	.83 (.38)
48. How to recheck answers	.87 (.34)	.90 (.32)	.82 (.39)	.85 (.36)
Testwiseness				
33. Taught guessing strategies	.71 (.46)	.80 (.42)	1.00 (.00)	.80 (.40)
44. Hints & strategies about multiple-choice type items	.71 (.46)	.70 (.48)	.68 (.48)	.69 (.47)
49. Clues to find the right answer	.70 (.46)	.70 (.48)	.86 (.35)	.74 (.44)

Note: Figures are rounded to two decimal places.

Table 4

Proportion (sd) of Teachers Engaging in Motivational Test Preparation Practices

Item number and question	Teaching Assignment-Grade 3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
21. Prizes or Incentives for preparing hard	.42 (.50)	.60 (.52)	.48 (.51)	.46 (.50)
35. Fun activity at conclusion of testing	.87 (.34)	.60 (.52)	.82 (.39)	.82 (.39)
43. Class contests to motivate students before testing	.39 (.50)	.50 (.53)	.45 (.51)	.42 (.50)
50. Created exciting classroom atmosphere with posters, etc	.47 (.51)	.50 (.53)	.77 (.43)	.56 (.50)
52. Rewards for good attendance during test week	.50 (.51)	.60 (.52)	.45 (.51)	.49 (.50)

Note: Figures are rounded to two decimal places.

Table 5

Proportion (sd) of Teachers Engaging in Coaching and Test Familiarization Test Preparation Practices

	Teaching Assignment-Grade			
Practice Item number and question	3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
Intensive Practice and Coaching				
22. How to use deductive reasoning skills	.95 (.23)	.90 (.32)	1.00 (.00)	.96 (.20)
30. Used *Scoring High*	.87 (.34)	1.00 (.00)	.69 (.52)	.93 (.26)
36. Special reviews & content drills	.76 (.43)	.70 (.48)	.73 (.46)	.73 (.45)
46. Focused instruction on extensive drills on items similar to that tested	.65 (.48)	.60 (.52)	.77 (.43)	.67 (.47)
Test Familiarization				
28. Used other tests of similar format to practice ITBS format	.87 (.34)	.80 (.42)	.95 (.21)	.89 (.32)
34. Practice questions like those found on ITBS	.71 (.46)	.70 (.48)	.64 (.49)	.68 (.47)
40. Showed students past versions of ITBS so they would know what to expect	.34 (.48)	.50 (.53)	.36 (.49)	.37 (.49)

Note: Figures are rounded to two decimal places.

Table 6

Proportion of Teachers Reporting Utilization of Preparation for Mathematics ITBS Testing

Item number and question	Teaching Assignment-Grade 3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
56. In general, I approach test preparation for the ITBS MATH tests by providing:				
A short test-taking orientation and brief practice sessions	.24	.30	.04	.18
A long, intensive coaching program that includes concentrated drill and practice on a sample of items or skills similar to those tested on the ITBS.	.26	.10	.35	.27
A broad-based approach with instruction to improve many skills but not necessarily any one skill tested on the ITBS.	.50	.60	.48	.52
I did not provide any test preparation for my students on the ITBS, other than day-to-day instruction for the regular curriculum.	.00	.00	.09	.03

Note: Figures reported have been rounded to two decimal places.

Table 7

Proportion of Teachers Reporting Utilization of Preparation for Reading ITBS Testing

Item number and question	Teaching Assignment-Grade 3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
56. In general, I approach test preparation for the ITBS READING tests by providing:				
A short test-taking orientation and brief practice sessions	.24	.20	.05	.17
A long, intensive coaching program that includes concentrated drill and practice on a sample of items or skills similar to those tested on the ITBS.	.26	.00	.27	.23
A broad-based approach with instruction to improve many skills but not necessarily any one skill tested on the ITBS.	.47	.70	.59	.55
I did not provide any test preparation for my students on the ITBS, other than day-to-day instruction for the regular curriculum.	.03	.10	.09	.06

Note: Figures reported have been rounded to two decimal places.

Table 8

Proportion (sd) of Teachers Reporting Utilization of Inappropriate Testing Practices

Item number and question	Teaching Assignment-Grade 3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
24. Used prior year test questions as practice	.43 (.50)	.60 (.52)	.41 (.50)	.44 (.50)
27. Recoded answer sheet	.14 (.35)	.50 (.53)	.38 (1.12)	.26 (.70)
29. Assigned test-related homework on weekends and vacations	.32 (.47)	.50 (.53)	.41 (.50)	.37 (.49)
31. Gave students as much time as needed so every student got to complete the test	.18 (.39)	.50 (.53)	.23 (.43)	.24 (.43)
38. Taught students how to answer an item thought to be on test	.62 (.49)	.60 (.52)	.59 (.50)	.60 (.49)
42. During test, gave hint or clue to the answer	.19 (.40)	.50 (.53)	.18 (.39)	.23 (.42)
45. During test, praised students who answered an item correct	.21 (.41)	.50 (.53)	.14 (.35)	.23 (.47)
47. Modified time schedule	.45 (.50)	.50 (.53)	.32 (.48)	.41 (.50)
51. Practice on items off current test with a change in stem or distractors	.21 (.41)	.60 (.52)	.14 (.35)	.24 (.43)
53. Encouraged best students to do their best	.18 (.39)	.40 (.52)	.09 (.29)	.18 (.39)
54. Practice on items off the current test	.11 (.31)	.60 (.52)	.09 (.29)	.17 (.38)

Note: Figures have been rounded to two decimal places.

Table 9

Ratings of Agreement (sd)with Statements Regarding Pressure to Improve Student Test Scores

	Teaching Assignment-Grade			
Pressure Item number and question	3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
Perception of Pressure [1]				
58. Encouraged by superiors to use any method to improve scores	2.61 (1.10)	2.50 (.97)	2.50 (.91)	2.54 (1.03)
59. Pressure from others to raise math scores	3.00 (1.04)	3.10 (.57)	2.82 (1.05)	2.94 (.98)
60. Principal pressured me spend increased time	2.76 (1.10)	3.00 (.67)	2.64 (1.09)	2.7 (1.04)
61. Pressure from others to raise reading scores	3.03 (1.05)	3.40 (.52)	3.00 (.98)	3.07 (.96)
62. Too much emphasis on test scores in my school	3.39 (.86)	3.30 (.67)	3.22 (.67)	3.33 (.77)
63. Aware of teachers or principals who have cheated to get higher scores	2.41 (1.21)	2.30 (1.34)	2.43 (1.12)	2.38 (1.19)
Intensity of Pressure [2]				
64. How much pressure to improve reading scores	2.37 (1.30)	2.50 (1.08)	2.96 (1.16)	2.60 (1.24)
65. How much pressure to improve math scores	2.47 (1.31)	2.50 (1.08)	2.96 (1.16)	2.64 (1.23)

[1] Items scaled 1= strongly disgree to 4=strongly agree. Rescaled for analyses 1=strongly disagree 4=strongly agree.

[2] Items scaled 1=No pressure to 5=Extreme pressure. Response options 2, 3, and 4 were unlabeled for respondents.

Table 10

Ratings of Agreement (sd) with Statements Regarding Value and Benefit of Testing Program

Item number and question	Teaching Assignment-Grade 3 (n=38)	4 (n=10)	5 (n=31)	Total (N=79)
66. Testing program valuable to me as a teacher	1.95 (.87)	2.20 (.63)	2.45 (.86)	2.41 (.85)
67. Testing beneficial for students	1.82 (.73)	2.10 (.99)	2.18 (.85)	1.97 (.81)
68. Tests provide me with information not obtained from other sources	1.82 (.93)	2.30 (.67)	2.00 (.82)	1.94 (.86)
69. Classroom performance an important indicator of my instructional effectiveness	1.76 (.82)	2.10 (1.29)	2.00 (1.02)	1.89 (.96)
70. Costs are justified given information gained from testing	1.74 (.79)	1.90 (.88)	2.14 (1.04)	1.87 (.89)

Note: Figures are rounded to the nearest whole number. Items originally scaled 1=strongly agree to 4=strongly disagree. Items were recoded for ease of presentation in table to 1=strongly disagree to 4=strongly agree.

Table 11

Relationship of Pressure to Improve Test Scores, Perceived Value and Benefit of Testing with Utilization of Test Information for Pupil Monitoring and Feedback

Item number & Activity[1]	Pressure[2] Q64 Reading	Pressure[2] Q65 Math	Value and Benefit[3] Q66 Valuable to teacher	Q67 Benefit to students	Q68 Provides unique info	Q69 Class Perform	Q70 Costs Justified
2. Informed pupils of status	.02	-.02	.05	.07	-.14	-.06	-.10
3. Informed parents of status	-.22	-.17	-.06	-.07	-.17	.04	-.12
4. Group pupils for instruction	-.18	-.20	-.05	.10	.07	.09	.03
5. Diagnosed learning difficulties	.14	.09	.02	.16	.21	.11	.18
6. Found pupils instructional level	.05	.00	.05	.21	.14	.13	.20
7. Assessed whether pupils were meeting district expectations	.04	.05	.17	.12	.11	.02	.02
8. Assigned students for remedial instruction	-.09	-.09	-.15	-.22	-.24*	.02	.06
9. Identified students for enrichment opportunities	-.32**	-.33**	-.04	-.16	-.13	-.15	-.17
10. Worked with counselors about an individual student	.16	.14	.22	.30*	.22	.27*	.29*

Note: Figures have been rounded to two decimal places. [1] Scaled 0=did not do, 1=did do. [2] Scaled 1=no pressure to 5=extreme pressure. [3] Scaled 1=strongly agree to 4=strongly disagree. * p≤.05, ** p≤.01.

Table 12

Relationship of Pressure to Improve Test Scores, Perceived Value and Benefit of Testing with Utilization of Test Information for Instructional and Curricular Decision-Making

	Pressure[2]		Value and Benefit[3]				
Item number & Activity[1]	Q64 Reading	Q65 Math	Q66 Valuable to teacher	Q67 Benefit to students	Q68 Provides unique info	Q69 Class Perform	Q70 Costs Justified
11. Evaluated instructional methods	.17	.18	.04	.06	-.02	.09	.03
12. Assessed teaching effectiveness	.27*	.24*	.11	.11	.09	.01	.12
13. Matched instructional methods to ability of students	.27*	.25*	.16	.22	.03	.15	.16
14. Revised scope & sequence	.27*	.29*	.27	.26*	-.04	.22	.19
15. Added lessons to curriculum	-.01	.04	.05	.15	-.08	.14	.14
16. Increased emphasis of ITBS content in curriculum	.22	.23	.13	.10	.09	.06	.07
17. Studied ITBS to assess match with instructional objectives	.02	.07	.17	.04	.02	.03	-.07
18. Increased time preparing for ITBS	-.07	-.06	.05	-.02	-.08	-.05	-.05
19. Eliminated existing topics	.01	.01	.01	.05	.07	-.02	.02

Note: Figures have been rounded to two decimal places. [1] Scaled 0=did not do, 1=did do. [2] Scaled 1=no pressure to 5=extreme pressure. [3] Scaled 1=strongly agree to 4=strongly disagree. * p≤.05, ** p≤.01.

Table 13

Relationship of Pressure to Improve Test Scores, Perceived Value and Benefit of Testing with Utilization of Inappropriate Testing Practices

	Pressure[2]		Value and Benefit[3]				
	Q64	Q65	Q66	Q67	Q68	Q69	Q70
Item Number and Description of Inappropriate Activity[1]	Reading	Math	Valuable to teacher	Benefit to students	Provides unique info	Class Perform	Costs Justified
24. Used prior year test questions as practice	.04	.05	.28*	.34**	.18	.15	.29*
27. Recoded answer sheet	.02	.02	.26*	.17	.25*	.25*	.12
29. Assigned test-related homework on weekends and vacations	.10	.10	.17	.23	-.04	.08	.09
31. Gave students as much time as needed so every student got to complete the test	.04	.05	.16	.20	.17	.15	.13
38. Taught students how to answer an item thought on test	.-.04	-.01	.03	.07	.12	.18	.01

Note: Figures have been rounded to two decimal places. [1] Scaled 0=did not do, 1=did do. [2] Scaled 1=no pressure to 5=extreme pressure. [3] Scaled 1=strongly agree to 4=strongly disagree.

Table 13 (continued)

Relationship of Pressure to Improve Test Scores, Perceived Value and Benefit of Testing with Utilization of Inappropriate Testing Practices

Item Number and Description of Inappropriate Activity[1]	Pressure[2] Q64 Reading	Pressure[2] Q65 Math	Value and Benefit[3] Q66 Valuable to teacher	Q67 Benefit to students	Q68 Provides unique info	Q69 Class Perform	Q70 Costs Justified
42. During test, gave hint or clue to the answer	.02	.03	.24*	.20	.17	.17	.10
45. During test, praised students with correct answer	.10	.11	.20	.24*	.17	.19	.17
47. Modified time schedule	.08	.10	.19	.30*	.21	.28*	.30*
51. Practice on items off current test with change	.05	.06	.27*	.24*	.21	.19	.13
53. Encouraged best students to do their best	.04	.08	.02	.26*	.08	.19	.12
54. Practice on items off the current test	.09	.11	.29*	.31**	.22	.19	.16

Note: Figures have been rounded to two decimal places. [1] Scaled 0=did not do, 1=did do. [2] Scaled 1=no pressure to 5=extreme pressure. [3] Scaled 1=strongly agree to 4=strongly disagree